DOCUMENT RESUME

ED 226 068 TM 830 127

AUTHOR Bennett, Richard P.
TITLE The Absolute Differential Area Technique for Testing Distributional Normality.
PUB DATE 28 Jan 83
NOTE 38p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 27-29, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; *Hypothesis Testing; *Statistical Distributions
IDENTIFIERS Chi Square Test; *Kolmogorov Smirnov Two Sample Test; Lilliefors Statistic; *Normal Distribution; Power (Statistics); Sample Size

ABSTRACT

The results of a study of find alternative techniques for testing distributional normality are presented. A group of statistical techniques--some established and some new--were compared using empirical techniques. One new technique which appears to have higher power than the Lilliefors test was subjected to a better definition. Distributions under the null hypothesis are presented for small sample sizes. Modelled relationships from these to the chi-square distributions for similar degrees-of-freedom are discussed. A final comparison to the Lilliefors test using the theoretical distributions indicated that the original findings were not supported. The Lilliefors test has the higher power. The new test is proposed as a potential alternative to the Kolmogorov-Smirnov test for general distributions. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED226068

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.P. Bennett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE ABSOLUTE DIFFERENTIAL AREA TECHNIQUE FOR TESTING DISTRIBUTIONAL NORMALITY

RICHARD P. BENNETT, Ed.D

TEXAS WOMAN'S UNIVERSITY

TM 830 127

THE ABSOLUTE DIFFERENTIAL AREA TECHNIQUE FOR TESTING DISTRIBUTIONAL NORMALITY

Summary

The results of a study to find alternative techniques for testing distributional normality are presented. A group of statistical techniques--some established and some new--were compared using empirical techniques. One new technique which appears to have higher power than the Lilliefors test was subjected to a better definition. Distributions under the null hypothesis are presented for small sample sizes. Modelled relationships from these to the chi-square distributions for similar degrees-of-freedom are discussed. A final comparison to the Lilliefors test using the theoretical distributions indicated that the original findings were not supported. The Lilliefors test has the higher power. The new test is proposed as a potential alternative to the Kolmogorov-Smirnov test for general distributions.

Introduction

Research results in the behavioral sciences are only as good as the power of the statistics used to achieve those results. If a new statistic can be developed that conducts a particular test with greater power than was possible by an earlier statistic, then behavioral science as a whole benefits. Distributional normality is an assumption which underlies many statistical procedures. Most researchers are content to accept the assumption of normality or to rely on the robustness of their statistics rather than to test the assumption, despite the fact that distributional normality can be tested in a variety of ways. Admittedly, some of these techniques are relatively complex for persons who do not possess a strong mathematical background. But others, which can be classified as non-parametric are relatively simple to use and understand. If the majority of users are more likely to use one of the non-parametric tests (if they use any), it is important that the tests have the greatest possible power for detecting non-normality.

Arising from the need for statistics with the highest possible power, the purposes of the present study are to: (1) examine three existing non-parametric techniques which can be used for testing distributional normality; (2) to evaluate the possibility of modifying two existing techniques which presently are not directly applicable to the

problem; and (3) to test a set of new techniques derived for this purpose.

To accomplish these purposes, the study used three phases. In the first, the various statistics were evaluated relative to the Lilliefors (1967) statistic. The latter statistic is a modification of the Kolmogorov-Smirnov test, specifically designed for testing distributional normality (Marascuilo & McSweeney, 1977). Those candidate statistics which demonstrated higher estimated power than the Lilliefors test were determined.

In the second phase, one statistic (the Absolute Differential Area Statistic or ADAS) was selected. Sampling distributions based on small sample sizes were developed for this statistic. Reevaluation of some of the data using the actual sampling distributions suggested that the ADAS was not as powerful as first thought as a test of distributional normality. However, since the ADAS can be used for more general tests, the study was continued. In the third phase, initial attempts were made to relate the ADAS to the Chi-square distribution.

Although the present study has yet to find a more powerful statistical test of distributional normality, the initial steps have been taken in defining a new statistic which has possible applications to general tests of distributional shape.

Background

Testing of distributional normality has been a concern of statisticians for many years. Many procedures have been proposed. Some of these have been relatively complex to use. Others have been simpler, often being of a non-parametric nature. A few of these will be reviewed briefly.

In 1947, Geary proposed a technique involving logarithms and a heavy use of moments. One step of the procedure involved the ratio of the sum of the absolute deviation scores to the square root of the sum of squares of the deviation scores. This procedure could only be used for normality testing. Anderson and Darling (1954) developed a test that could be used with any specified distribution. However, their technique involved some knowledge of calculus. While this knowledge might be available to statisticians, it is not generally found in users, at least those in the behavioral sciences.

The test suggested by Shapiro and Wilk (1965) used an analysis of variance approach. The test was somewhat complex in that it used matrix and vector notation. Again, these are methods only acquired with a wider background in mathematics than that obtained by many users.

The Chi-square goodness-of-fit test and the Kolmogorov-Smirnov single sample test (Siegel, 1954) are non-parametric tests which are applicable to any distribution that can be divided into intervals. These tests were examined by Durbin

(1961). He concluded that the nature of the Kolmogorov-Smirnov test made it the more desirable of the two since it treats data at the ordinal level of measurement while the Chi-square treats the same data as nominal. Even so, the Chi-square test showed better power that the Kolmogorov-Smirnov under certain conditions. Neither test was examined specifically for the normal distribution.

Although the Kolmogorov-Smirnov test treats the data as ordinal, it establishes a statistic based on a single point in the distribution. This might explain its occasional poor performance relative to the Chi-square. To gain more power relative to testing distributional normality, Lilliefors (1967) proposed a modification of the Kolmogorov-Smirnov test using a specific scheme for establishing the intervals of the distribution, and a sampling distribution derived for use only with the normality test. This test was specifically identified with respect to normality by Marascuilo and McSweeney (1977) in their very thorough presentation of non-parametric techniques. More recently, d'Agostino and Cureton (1972) suggested a test of normality in which the skew of the sample is converted to a t-statistic using a rather complex but strictly algebraic formula.

By using all data, the skew test would appear to overcome one of the questionable aspects of both the Kolmogorov-Smirnov and the Lilliefors tests, the use of the data from

a single interval to stand for the whole distribution. It was this consideration that led to the definition of the various alternative statistics described below. There was an instinctive feeling that a test which examines all intervals within a distribution would have advantages over one which considered only one interval.

Description of Statistics Tested

The statistical tests evaluated by this study all have one common characteristic. The first step taken consists of dividing the standarized normal distribution into areas for which expected frequencies can be defined and into which observed standarized scores can be sorted. Two general schemes are used based on equal areas or on equal intervals. Equal area procedures involve dividing the normal distribution into a selected number of equal areas for which the boundaries are defined. All areas have the same expected frequency. For this study, up to five different schemes were used dependent on sample size. The five involved dividing the normal distribution into 5, 10, 20, 40 or 80 equal areas. A particular scheme was not used if the number of equal areas selected exceeded the sample size. For example, tests on samples of 40 scores did not include dividing the normal distribution into 80 equal areas.

The equal interval schemes involved dividing the normal distribution into unequal areas of equal width. In this

study, the number of intervals selected were 7 (1.0 standard deviation in width) or 13 (0.5 standard deviation in width). In each case, the middle interval was centered at $z = 0.0$ and the two outer intervals were extended to include all outlying scores. The test involving 13 equal intervals was not used with samples of 10 scores.

The criterion test for the study was the Lilliefors (1967) statistic, a modification of the Kolmogorov-Smirnov test specifically for normal distributions. This test is accomplished by dividing the normal distribution into n equal areas where n is the size of the sample to be tested. Expected frequencies are one per interval. Cumulative proportional frequency distributions are then constructed for the observed scores and for the expected scores. The greatest absolute difference between the two is the Lilliefors statistic which is compared to established tables of critical values (Lilliefors, 1967).

The established statistics evaluated in this study were the Chi-square test and the Kolmogorov-Smirnov test. While the Lilliefors test specifies a distribution based on n intervals, these two statistics permit the choice of intervals to be arbitrary. To test the effects of the choice of interval scheme, the various distributional schemes described above were employed. Briefly, the Chi-square statistic compares the expected and observed frequencies while the

Kolmogorov-Smirnov test determines the greatest absolute difference between observed and expected cumulative proportional frequency distributions.

Two non-parametric statistical tests were used in a way for which the established tables of critical values were not appropriate. The distributional schemes described above were not used. The Mann-Whitney U-test assumes that two sets of independently selected scores exist. The two sets are ranked as a single group then separated. If the two sets of scores represent a common population, their ranks will be randomly assigned and equal mean ranks should result. To use the Mann-Whitney U-test, one of the two sets of scores was assigned to represent the normal distribution. Two different methods were used. In the first, the normal distribution was divided into n equal areas, where n was the sample size. The medians of the n areas served as one of the two sets of data. The second technique for locating the n scores divided the normal distribution into $n + 1$ equal areas. The resulting boundaries were used as one set of scores. With either technique, the hypothesis was that the n randomly chosen scores would mingle with the n chosen arbitrarily and result in no difference in mean rank.

The second non-parametric test which was used in an alternative manner was the Wilcoxon test. Two sets of scores, matched in pairs, are defined. Differences between pairs are computed and ranked without reference to the sign

of the difference. The rank sum of the positive differences is compared to the rank sum of the negative ones to test the assumption that the numbers and magnitudes of the two sets of differences are equal within statistical significance. If the original two groups of scores are similar, the magnitude and direction of the differences will be randomly distributed and no difference in the sum of the ranks will occur. The two distributional schemes employed with the Mann-Whitney U-test were also used with the Wilcoxon. Pairing of the two sets of scores was accomplished by rank order. The highest score of the randomly selected set of n scores was paired with the highest arbitrary score, the second highest randomly chosen score to the second highest arbitrary score, and so on.

The new statistics tested were all loosely based on the Kolmogorov-Smirnov approach. But where the Kolmogorov-Smirnov (and Lilliefors) considered only the interval having the largest difference between the proportional cumulative frequency distributions, the new statistics were designed to consider the pattern of differences in all intervals. All of these statistics involve establishing cumulative frequency histograms or polygons for both expected and observed frequencies. Distributional schemes used with the Chi-square and Kolmogorov-Smirnov tests were employed here also.

The Absolute Differential Area statistic is computed by summing the magnitudes of the differences between the expected and observed cumulative frequency histograms. The resulting statistic must be positive, suggesting that the sampling distribution should be skewed. The Algebraic Differential Area statistic sums the differences algebraically. The sampling distribution for this statistic should be summetrical with a mean of zero. The Squared Differential Area Statistic sums the squares of the differences between the two cumulative frequency histograms. Like the Absolute Differential Area statistic, it must be positive, therefore a skewed sampling distribution can be predicted for this statistic. The Absolute Differential Polygon Statistic is identical to the Absolute Differential statistic with the exception that the cumulative frequency polygons were used in place of the histograms. Again, a skewed sampling distribution seems probable. (An Algebraic Differential Polygon statistic produces values identical to the Algebraic Differential Area Statistic.) Finally, as an extension of the Squared Differential Area statistic, a Differential Deviation statistic was developed in which the standard deviation was computed for the differences between the two cumulative frequency histograms.

Methodology

The population for the study consisted of one million normally distributed standardized (z) scores. Selection

of samples was accomplished using the Rand Corporation (1955) table of random numbers. One hundred samples of six digits were drawn in each of four sample sizes: n = 10, 20, 40, and 80. Starting points within the table were randomly selected for each of the 400 samples.

Each sample was first standardized, then tested for normality using the Lilliefors (1967) test which served as the criterion for the study. Each sample was further tested using a battery of proposed tests. These included existing non-parametric tests which were not considered optimum for this task (the Chi-square and Kolmogorov-Smirnov tests), existing non-parametric tests being used in an alternate way (Mann-Whitney U-test and Wilcoxon test), and a group of new statistics based loosely on the Kolmogorov-Smirnov procedure. Some techniques such as the Chi-square were tested under various distributional schemes to determine if there were any effects due to the method of conducting the test.

The resulting data consisted of 100 scores in each of four sample sizes for each test or combination of test and distributional scheme employed. These scores were tabulated for use as the base distribution under the true null hypothesis. For the Lilliefors, and the various options of the Chi-square and Kolmogorov-Smirnov tests, these merely served as checks against pre-established statistical distributions. For all other tests these distributions were used to establish a rough estimate of a critical value for an alpha of

.05. For symmetrical distributions, this was accomplished by computing the mean and standard error, and presuming the normality of the sampling distribution. This critical value was compared to the five extreme scores of the distribution as an estimate of alpha. For skewed distributions the boundary separating the five extreme scores from the rest of the distribution was used as the critical value.

The next step involved obtaining samples from a non-normal distribution. To obtain a skewed population, one need only subject each element of the original normal population to a power formula. However, it was sufficient in this instance to subject only the elements of the selected samples to the formula. The sample was then standardized for normality testing (as were the original samples). For simplicity, and to eliminate differences between samples, the 400 original samples from the normal distribution were reused after skewing.

Each skewed sample was tested for normality by the full battery of proposed statistics. For any specific test (or combination of test and distributional scheme) and sample size, the power of the test was defined as the proportion of the 100 samples whose normality was rejected. The existing tables of critical values were used where appropriate. The critical values established by the first set of tests were used otherwise. The estimation of power

for the various tests was accomplished at two different values of skew, 0.92 and 1.9.

Results of the power evaluations were examined relative to the power demonstrated by the Lilliefors test on the same set of samples. As a result, the Absolute Differential Area statistic (ADAS) was chosen for more precise definition of distributional characteristics under the null hypothesis. This consisted of establishing exact sampling distributions for sample sizes ranging from $n = 3$ to $n = 10$. These sampling distributions were then related to the appropriate Chi-square distributions by entering equivalent critical values for both into a modeling program. The results of the modeling provided an initial algebraic relationship intended to transform an ADAS value into a Chi-square for statistical evaluation.

Results

The first phase of the project consisted of determining power estimates for the statistics under consideration for the various sample sizes and distributional schemes described. In each instance, the Lilliefors power figures are presented to serve as a criterion.

The power of the Chi-square statistic under various configurations is presented in Table 1. The bulk of the table confirms the view that the Chi-square would be a less powerful statistic. The data does show that greater

deviations from normality tested with larger sample sizes and using the distributional scheme of 13 equal width intervals would result in a Chi-square test with power approximating that of the Lilliefors test.

The performance of the Kolmogorov-Smirnov statistic (Table 2) clearly shows its inferiority to the Lilliefors statistic. Although the Kolmogorov-Smirnov was intended for use with any established distribution, its performance under these conditions could lead one to have some dobts as to its power relative to other non-normal distributions.

In comparison, the estimated power figures for the modified Mann-Whitney U-test (Table 3) equal or exceed those of the Lilliefors test as all sample sizes. Caution must be exercised, however, since the estimated power figures are based on critical values which were empirically established from a limited set of 100 samples. The estimated power figures from the modified Wilcoxon test which are also presented in Table 3, are far less impressive. Any further consideration for the purpose of testing normality would seem inappropriate.

The first of the new statistics is the Absolute Differential Area statistic. Its estimated performance is presented in Table 4. Like the Modified Mann-Whitney, this statistic demonstrated power estimates equal to or better than the Lilliefors test. Of particular importance

is the power shown by the new statistic using the n equal area distribution scheme. For moderate sample sizes (n = 20 and 40), the new statistic appears to be significantly more powerful.

In contrast, the Algebraic Differential Area statistic was comparable to the Absolute Differential Area statistic for distributional schemes using equal areas, but performed poorly when the distributional scheme was one of equal interval width. This data is presented in Table 5.

Performance of the Squared Differential Area statistic (Table 6) was equal to or better than the Lilliefors statistic under all distributional schemes. Power figures demonstrated by this statistic were very comparable to those of the Absolute Differential Area statistic (Table 4). The same conclusion could be made about the performance of the Absolute Differential Polygon statistic (Table 7). This is to be expected given the similarity of the latter statistic (based on cumulative frequency polygons) to the Absolute Differential Area statistic (based on cumulative frequency histograms).

Finally, performance of the Differential Deviation statistic (Table 8) was analogous to that of the Squared Differential Area Statistic (Table 6). Since these are the standard deviation and the raw score sum of squares for a set of scores, the analogy is also to be expected.

The development of the performance figures for the various statistics marked the end of the first phase of the study. At this point a decision was necessary regarding the statistic to be selected for further examination. The Mann-Whitney U-test was set aside because it represents a questionable use of an established statistic. It remains available for future consideration. It seemed more appropriate to focus attention on a further modification of a procedure that is generally considered to be designed for the task of defining distributional shape, the Kolmogorov-Smirnov. Of the five new statistics, the Algebraic Differential Area statistic was eliminated because of its poor performance on the equal interval distributional schemes. The remaining four were very comparable in their estimated power performance, so selection among them was more of an arbitrary choice than one based on any particular rationale. None of the four could be expected to produce a symmetrical sampling distribution. Since the initial effort on the sampling distributions was accomplished without the aid of a computer, the Absolute Differential Area statistic (ADAS) was chosen for its ease in computation.

The second phase of the study consisted of theoretically defining the sampling distribution for the ADAS. Some limitations had to be imposed given the wide variety of distributional schemes (equal area and equal interval

width) employed in the first phase, plus the potential for additional schemes. The decision was made to focus specifically on the distributional scheme of n equal areas where n is the sample size. This scheme involves an expected frequency of 1 for each interval defined in the distribution. The cumulative frequency distribution becomes a stair-step with the cumulative frequency at any step being equal to the number of the interval in the distribution counting from the lowest interval.

For a given sample size (and number of intervals), establishing the sampling distribution consisted of determining the number of unique ways the n observed elements could be divided among the n intervals, then computing the ADAS for each permutation of each combination. As an example, consider the process for a sample size of three elements. There are three intervals for this sample size, each having an expected frequency of one element. There are three distinct combinations of observed frequencies possible: (1) all three scores in one interval; (2) two scores in one interval and one in another; and (3) one score in each interval. The first of these combinations has three permutations in which the three scores are assigned to each of the three intervals in turn. The second combination has six permutations based on three choices for assigning the group of two scores coupled with two subse-

quent choices for the single score. The third combination has only the one permutation.

General determination of the number of permutations is accomplished using combinatorial theory. The general equation is:

$$\binom{n}{a_0\ a_1\ \ldots\ a_n} = \frac{n!}{a_0!a_1!\ldots a_n!}$$

where: a_i is the number of times the ith frequency occurs in the combination

Returning to the example, for the first combination (3-0-0), the frequency 0 occurs twice, the frequencies 1 and 2 do not occur and the frequency 3 occurs once. The number of permutations will be:

$$\binom{3}{2\ 0\ 0\ 1} = \frac{3!}{2!0!0!1!} = 3$$

For the second combination (2,1,0), the number of permutations is:

$$\binom{3}{1\ 1\ 1\ 0} = \frac{3!}{1!1!1!0!} = 6$$

Each of these permutations or allocation of frequencies forms a member of the set of possible observed frequency distributions. For each, an observed cumulative frequency distribution was defined and the ADAS computed. As an example, take one of the permutations of the first combination: 0, 3, 0. The observed cumulative frequency distri-

bution is 0, 3, 3 which is compared to the expected cumulative frequency distribution, 1, 2, 3. The differences between the two by interval are -1, +1, 0. Summing their absolute values produces an ADAS equal to 2 for this permutation.

Coupled with the ADAS for a particular permutation is the probability of its occurrence. These are determined from the multinomial distribution using the general equation:

$$\binom{n}{b_1\ b_2\ \ldots\ b_J}\left(\frac{1}{n}\right)^n$$

where: b_j is the number of scores in the jth interval This simplified equation can be used because each interval has the same probability of receiving a score. For the example used above the probability of 0, 3, 0 is:

$$\binom{3}{0\ 3\ 0}\left(\frac{1}{3}\right)^3 = \frac{3!}{0!3!0!}\left(\frac{1}{3}\right)^3 = 1/27$$

Since the probabilities are the same for all permutations of a particular combination, the probability of 3, 0, 0 and 0, 0, 3 are also 1/27.

The same value for the ADAS may appear in different permutations associated with different combinations. To determine the probability of a particular value of ADAS, one must add the probabilities of all permutations that produce the specified value. For example, the permutation

0, 3, 0 from the first combination and the permutations 2, 1, 0 and 0, 1, 2 from the second combination all produce an ADAS of 2. The probability of the first is 1/27, while it is 3/27 for each of the latter two. This results in a total probability of 7/27 that a sample of three scores will give an ADAS of 2. Similar probabilities can be derived for the full range of ADAS values for a given sample size. For n - 3, the ADAS can range from 0 to 3 with probabilities as follows: $p(0) = 6/27$; $p(1) = 12/27$; $p(2) = 7/27$; $p(3) = 2/27$. From this probability distribution, the centile points for significance testing are derived.

Based on the procedure just described, Fortran programs were written for the DEC20 computer. Using these, sampling distributions were defined for sample sizes ranging from 3 to 10. For reasons which will be discussed below, these were modified by dividing each distribution by the square root of the sample size. Specific centile points from the resulting probability distributions are presented in Table 9.

As indicated before, the Lilliefors test specifies a distributional scheme using n equal sample sizes. The initial development of sampling distributions for the ADAS followed the same scheme. However, before confining the ADAS to this scheme, several sampling distributions were examined for cases in which the sample size was an

integral product of the number of intervals. For example, sampling distributions were defined for samples of 6 and 9 scores distributed among 3 intervals. Similar distributions were derived for 4 and 5 interval situations. The procedures used required only small modifications from those described above. The result of defining these additional sampling distribution was to show that for a given number of equal area intervals, all sampling distributions could be related to one common distribution by the simple step of dividing the sampling distribution by the square root of the sample size. This means that the sampling distribution of three scores distributed among three intervals when divided by the square root of 3 is the same as that of six scores distributed among three intervals when divided by the square root of 6. These comparisons were not perfect due to the discontinuities that exist in the sampling distributions. However, equality to three and four place accuracy was experienced. Thus it seems reasonable to conclude that larger sample sizes would fit the rule with greater accuracy. For this reason, the data presented in Table 9 represents the modified ADAS in which the sum of the absolute differences between the cumulative frequency histograms is divided by the square root of the sample size.

Having established the actual sampling distributions, it was possible to determine how well the actual power figures measured up to the estimated ones. The results

of this comparison are presented in Table 10. The loss in power that occurred was sufficient to raise serious questions regarding the value of the ADAS for testing normality. The data clearly indicates that the Lilliefors is the far better test insofar as this limited comparison permits. Nor does it seem likely that extensions of the sampling distribution will change this fact. However, the ADAS is a more general statistic than the Lilliefors since it can be used to test any expected distribution which can be divided into equal areas. Consequently, the decision was made to continue further definition of the ADAS.

While it was possible to continue the derivation of further sampling distributions in this manner, two characteristics of the ADAS distribution suggested an alternative approach: (1) its positively skewed nature; and (2) the increasing magnitude of the distribution with sample size. These offered the possibility that a mathematical formulation could tie the ADAS to the Chi-square distribution. To examine this possibility, certain centile points for the two distributions were supplied to the MLAB program (Knott & Reece, 1972) on the DEC20. To be more specific, centile points for an ADAS of n intervals was compared to those of a Chi-square with both n degrees of freedom and, again, with $n-1$ degrees of freedom. The latter was presumed to be the more logical choice but both were examined.

The results of the modeling phase of the study were as follows. The best relationship between one of the ADAS distributions and an appropriate Chi-square distribution was linear. Of the two Chi-square distributions modeled with the ADAS of $\underline{n}$ intervals, the Chi-square with $\underline{n}$-1 degrees of freedom was modeled with less error. The linear relationship established between the ADAS and the Chi-square varied with $\underline{n}$ but did so in a manner suggesting a trend.

The slopes of the eight modeled relationships ($\underline{n}$ = 3 to $\underline{n}$ = 10) were modeled against $\underline{n}$. The result was a new linear function of slope in terms of $\underline{n}$. The intercepts of the eight modeled relationships did not fare as well. The best fit to the limited data available was a quadratic function. These two functions provide a relationship between the ADAS and the Chi-square of the following form:

$$\text{Chi-square} = f_1(\underline{n})_{lin} \cdot \text{ADAS} + f_2(\underline{n})_{quad}$$

This result suggests the need for further extension of the ADAS sampling distribution for larger values of $\underline{n}$ to establish a more regular relationship between the ADAS and the Chi-square. Further, this finding is consistent with other non-parametric statistics which need specific sampling distributions with small sample sizes.

Discussion

The first conclusion to be drawn from this study is that the promise shown by the Absolute Differential Area

statistic (ADAS) was not borne out when the estimated power values were compared to the real power figures on a limited basis. Some shrinkage of power was to be expected consistent with the use of a small number (100) of samples for establishing the estimated critical values. Such a large shrinkage was not anticipated. The lesson to be drawn from this is that the estimates needed a much wider base, perhaps a magnitude higher, i.e. 1000 samples. From the actual power figures it is obvious that the ADAS will not challenge the Lilliefors statistic as a test for normality.

The ADAS, however, may find other uses. While it was defined with the express intention of testing normality, unlike the Lilliefors test it is not limited to this purpose. The ADAS is like the Kolmogorov-Smirnov test in that it appears to be applicable to any distribution given that the expected distribution can be divided into equal areas. Although the Kolmogorov-Smirnov test is designed for this purpose, the preeminence of the Lilliefors test for normality suggests the Kolmogorov-Smirnov may be less powerful for other purposes as well.

Although there may be other uses for ADAS, identifying them would seem to be a secondary priority for the moment. The first priority lies with establishing a more complete picture of the sampling distribution. This will consist of deriving sampling distributions for more than 10 inter-

vals until it is possible to establish at what point and in what way a function of the ADAS can be approximated as a Chi-square. Only then will it be practical to test the ADAS against the Kolmogorov-Smirnov test for distributions other than normal.

Bibliography

Anderson, T. W., & Darling, D. A. A test of goodness of fit. Journal of the American Statistical Association, 1943, 49, 765-769.

d'Agostino, R. B., & Cureton, E. E. Test of normality against skewed alternative. Psychological Bulletin, 1972, 78, 262-265.

Durbin, J. Some methods of constructing exact tests. Biometrika, 1961, 48, 41-55.

Geary, R. C. Testing for normality. Biometrika, 1947, 34, 209-242.

Knott, G. D., & Reece, O. K. MLAB: A civilized curve fitting system. Proceedings of the Online '72 International Conference, 1972, 1, 497-526.

Lilliefors, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, 1967, 62, 399-402.

Marascuilo, L. A., & McSweeney, M. Non-parametric and distribution free methods for the social sciences. Monterey, Cal.: Brook/Cole Publishing Company, 1977.

Rand Corporation. A million random digits with 100,000 normal deviates. Glencoe, Ill.: The Free Press, 1955.

Shapiro, S. S., Wilk, M. B. An analysis of variance test for normality (complete samples). Biometrika, 1965, 52, 591-611.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Company, Inc. 1956.

Table 1

Performance of the Chi-square Statistic

	Power[a]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Chi-square								
5 intervals	.02	.01	.05	.12	.09	.20	.33	.57
10 intervals	.06	.05	.11	.21	.03	.20	.64	.97
20 intervals	X	.04	.08	.24	X	.11	.46	.91
40 intervals	X	X	.07	.14	X	X	.37	.84
80 intervals	X	X	X	.19	X	X	X	.67
7 intervals	.02	.02	.14	.34	.12	.09	.47	.99[b]
13 intervals	X	.11	.07	.29	X	.38	.78	.96

[a]Power defined by established tables of critical values.

[b]Power exceeds corresponding Lilliefors power.

Table 2

Performance of the Kolmogorov-Smirnov Statistic

	Power[a]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Kolmogorov-Smirnov								
5 intervals	.00	.00	.00	.00	.00	.00	.02	.12
10 intervals	.00	.00	.00	.00	.00	.01	.02	.17
20 intervals	X	.00	.00	.00	X	.01	.05	.22
40 intervals	X	X	.01	.00	X	X	.07	.26
80 intervals	X	X	X	.00	X	X	X	.32
7 intervals	.00	.00	.00	.00	.00	.00	.00	.00
13 intervals	X	.00	.00	.00	X	.00	.00	.12

[a]Power defined by established tables of critical values.

Table 3

Performance of the Modified Statistics

	Power[a]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Mann-Whitney								
n intervals	.07	.22[b]	.66[b]	.78[b]	.31[b]	.61[b]	.96[b]	1.00[b]
n + 1 intervals	.08	.26[b]	.64[b]	.79[b]	.23[b]	.60[b]	.96[b]	1.00[b]
Wilcoxon								
n intervals	.03	.12	.17	.48	.05	.15	.26	.52
n + 1 intervals	.04	.05	.18	.43	.01	.06	.14	.52

[a]Lilliefors power defined by established tables of critical values; remaining power estimated from empirical data

[b]Power exceeds corresponding Lilliefors power.

Table 4

Performance of the Absolute Differential Area[a] Statistic

	Power[b]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Absolute Differential Area								
5 intervals	.07	.15	.25	.54[c]	.21	.40	.77	.97
10 intervals	.09[c]	.32[c]	.38[c]	.60[c]	.30[c]	.65[c]	.88[c]	.99[c]
20 intervals	X	.38[c]	.32[c]	.73[c]	X	.67[c]	.87[c]	.99[c]
40 intervals	X	X	.31[c]	.77[c]	X	X	.84[c]	.99[c]
80 intervals	X	X	X	.78[c]	X	X	X	1.00[c]
7 intervals	.14[c]	.16	.44[c]	.59[c]	.32[c]	.50[c]	.91[c]	1.00[c]
13 intervals	X	.28[c]	.41[c]	.76[c]	X	.68[c]	.95[c]	1.00[c]

[a]Absolute area difference between observed and expected cumulative frequency histograms.

[b]Lilliefors power defined by established tables of critical values; remaining power estimated from empirical data.

[c]Power exceeds corresponding Lilliefors power.

Table 5

Performance of the Algebraic Differential Area[a] Statistic

	Power[b]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Algebraic Differential Area								
5 intervals	.01	.12	.50[c]	.77[c]	.03	.26	.91[c]	1.00[c]
10 intervals	.05	.20[c]	.58[c]	.79[c]	.13	.53[c]	.89[c]	.99[c]
20 intervals	X	.32[c]	.56[c]	.82[c]	X	.65[c]	.97[c]	1.00[c]
40 intervals	X	X	.65[c]	.81[c]	X	X	.97	1.00[c]
80 intervals	X	X	X	.80[c]	X	X	X	1.00[c]
7 intervals	.00	.00	.02	.07	.00	.00	.11	.27
13 intervals	X	.05	.05	.05	X	.09	.16	.16

[a]Algebraic area differnce between observed and expected cumulative frequency histograms.

[b]Lilliefors power defined by established tables of critical values; remaining power estimated from empirical data.

[c]Power exceeds corresponding Lilliefors power.

Table 6

Performance of the Squared Differential Area[a] Statistic

	Power[b]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Squared Differential Area								
5 intervals	.11[c]	.22[c]	.31[c]	.65[c]	.29	.51[c]	.81[c]	.98[c]
10 intervals	.08	.31[c]	.33[c]	.64[c]	.26	.64[c]	.86[c]	.99[c]
20 intervals	X	.31[c]	.28[c]	.68[c]	X	.65[c]	.85[c]	.99[c]
40 intervals	X	X	.32[c]	.71[c]	X	X	.80	.99[c]
80 intervals	X	X	X	.71[c]	X	X	X	.99[c]
7 intervals	.06	.17	.33[c]	.49	.21	.45	.80	.98[c]
13 intervals	X	.21[c]	.35[c]	.68[c]	X	.63[c]	.95[c]	.99[c]

[a]Sum of squares of differences between observed and expected cumulative frequency histograms.

[b]Lilliefors power defined by established tables of critical values; remaining power estimated from empirical data.

[c]Power exceeds corresponding Lilliefors power.

Table 7

Performance of the Absolute Differential Polygon[a] Statistic

	Power[b]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Absolute Differential Polygon								
5 intervals	.07	.13	.31[c]	.67[c]	.24[c]	.39	.82[c]	.98[c]
10 intervals	.10[c]	.31[c]	.37[c]	.64[c]	.33[c]	.67[c]	.87[c]	.99[c]
20 intervals	X	.33[c]	.30[c]	.76[c]	X	.67[c]	.88[c]	.99[c]
40 intervals	X	X	.32[c]	.80[c]	X	X	.86[c]	.99[c]
80 intervals	X	X	X	.79[c]	X	X	X	1.00[c]
7 intervals	.03	.27[c]	.46[c]	.53[c]	.24[c]	.62[c]	.83[c]	1.00[c]
13 intervals	X	.25[c]	.41[c]	.80[c]	X	.70[c]	.94[c]	1.00[c]

[a]Absolute area difference between observed and expected cumulative frequency polygons.

[b]Lilliefors power defined by established tables of critical values; remaining power estimated from empirical data.

[c]Power exceeds corresponding Lilliefors power.

Table 8

Performance of the Differential Deviation[a] Statistic

	Power[b]							
	Skew = .92				Skew = 1.9			
Statistic	n = 10	20	40	80	10	20	40	80
Lilliefors	.08	.19	.26	.50	.21	.49	.80	.97
Differential Deviation								
5 intervals	.04	.17	.18	.27	.20	.42	.62	.81
10 intervals	.11[c]	.24[c]	.25	.55[c]	.27[c]	.58[c]	.85[c]	.99[c]
20 intervals	X	.27[c]	.30[c]	.63[c]	X	.59[c]	.85[c]	.99[c]
40 intervals	X	X	.26	.66[c]	X	X	.82[c]	.99[c]
80 intervals	X	X	X	.64[c]	X	X	X	.99[c]
7 intervals	.07	.17	.35[c]	.49	.21	.48	.82[c]	.98[c]
13 intervals	X	.20	.35[c]	.66[c]	X	.61[c]	.95[c]	.99[c]

[a]Standard deviation of stepwise differences between observed and expected cumulative frequency histograms.

[b]Lilliefors power defined by established tables of critical values; remaining power estimated from empirical data.

[c]Power exceeds corresponding Lilliefors power.

Table 9

Critical Values of the Absolute Differential Area Statistic

%tile	n = 3	4	5	6	7	8	9	10
1	-.455	-.098	.057	.193	.443	.581	.686	.781
2	-.154	-.006	.151	.322	.550	.752	.808	.974
5	-.056	.149	.344	.498	.775	.965	1.079	1.247
10	.094	.317	.501	.699	.968	1.192	1.346	1.504
20	.282	.511	.738	.938	1.242	1.502	1.688	1.880
30	.420	.667	.912	1.142	1.465	1.753	1.968	2.186
50	.666	.962	1.250	1.527	1.901	2.196	2.525	2.797
70	.935	1.287	1.656	2.003	2.450	2.869	3.230	3.577
80	1.118	1.532	1.944	2.343	2.842	3.321	3.738	4.624
90	1.378	1.880	2.382	2.880	3.466	4.031	4.544	5.028
95	1.622	2.194	2.788	3.360	4.033	4.683	5.270	5.843
98	1.854	2.572	3.254	3.921	4.652	5.451	6.118	6.798
99	2.027	2.795	3.567	4.308	5.159	5.969	6.682	7.444
99.9	2.341	3.405	4.392	5.361	6.418	7.406	8.265	9.213

Table 10

Comparative Performance of the Absolute Differential Area Statistic

Sample Size	No. of Intervals	Power			
		Skew = .92		Skew = 1.9	
		Estimated	Actual	Estimated	Actual
10	5	.07	.00	.21	.00
10	10	.09	.00	.30	.00
20	5	.15	.00	.40	.00
20	10	.32	.00	.65	.00
40	5	.25	.00	.77	.16
40	10	.38	.00	.88	.10
80	5	.54	.01	.97	.35
80	10	.60	.01	.99	.40